IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.x, No.x, July xxxx, pp. 1~5
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.xxxx

Electroencephalogram-Based Emotion Classification Using Machine Learning and Deep Learning Techniques

Gst. Ayu Vida Mastrika Giri*[1], Made Leo Radhitya[2]
[1]Program Studi Informatika, Fakultas MIPA, Universitas Udayana, Bali, Indonesia
[2]Program Studi Teknik Informatika, Institut Bisnis dan Teknologi Indonesia, Bali, Indonesia
e-mail: ***[1]vida@unud.ac.id**, [2] leo.radhitya@instiki.ac.id

Abstrak

Electroencephalogram (EEG) merekam aktivitas otak sebagai arus listrik untuk melihat emosi. Seiring dengan meningkatnya minat terhadap hubungan emosional antara manusia dan komputer, algoritma pengenalan emosi yang dapat diandalkan menjadi sangat penting. Penelitian ini mengklasifikasikan gelombang EEG menggunakan machine learning dan deep learning. Muse EEG headband dengan empat saluran merekam emosi netral, negatif, dan positif pada dataset Feeling Emotions EEG yang tersedia untuk umum. Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), dan Gated Recurrent Unit (GRU) digunakan untuk deep learning dalam penelitian ini, sementara SVM, K-NN, dan MLP digunakan untuk pembelajaran mesin. Model-model tersebut dievaluasi berdasarkan akurasi, presisi, recall, dan F1-Score. SVM, K-NN, dan MLP memiliki nilai akurasi sebesar 0.98, 0.95, dan 0.97. Metode deep learning CNN, LSTM, dan GRU memiliki akurasi 0,98, 0,82, dan 0,97. SVM dan CNN unggul dalam hal akurasi, presisi, recall, dan F1-Score. Penelitian ini menunjukkan bahwa machine learning dan deep learning dapat mengklasifikasikan sinyal EEG untuk mengidentifikasi emosi. Hasil akurasi yang tinggi, terutama dari SVM dan CNN, menunjukkan bahwa model-model ini dapat digunakan dalam sistem interaksi manusia-komputer yang sadar akan emosi. Penelitian ini menambah penelitian klasifikasi emosi berbasis EEG dengan mengungkapkan pemilihan model dan strategi penyesuaian parameter untuk klasifikasi yang lebih baik.

Kata kunci— *deep learning, electroencephalogram, emosi, klasifikasi, machine learning*

Abstract

Electroencephalogram (EEG) records brain activity as electrical currents to discern emotions. As interest in human-computer emotional connections rises, reliable and implementable emotion recognition algorithms are essential. This study classifies EEG waves using machine and deep learning. A four-channel Muse EEG headband recorded neutral, negative, and positive emotions for the publicly available Feeling Emotions EEG dataset. Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) were utilized for deep learning, while SVM, K-NN, and MLP were used for machine learning. The models were assessed for accuracy, precision, recall, and F1-Score. SVM, K-NN, and MLP have accuracy scores of 0.98, 0.95, and 0.97. Deep learning methods CNN, LSTM, and GRU had 0.98, 0.82, and 0.97 accuracy. SVM and CNN surpassed other approaches in accuracy, precision, recall, and F1-Score. The research shows that machine learning and deep learning can classify EEG signals to identify emotions. High accuracy results, especially from SVM and CNN, suggest these models could be used in emotion-aware human-computer interaction systems. This study adds to EEG-based emotion classification research by revealing model selection and parameter tweaking strategies for better categorization.

Keywords— *classification, deep learning, electroencephalogram, emotion, machine learning*

Received June 1st,2012; Revised June 25th, 2012; Accepted July 10th, 2012

1. INTRODUCTION

Neuroscience and computer science studies have intertwined. In recent years, neural networks and deep learning in neuroscience have garnered interest for their ability to analyze complicated neurological data using artificial intelligence methods like machine learning and deep learning. Today, commercial headsets can collect EEG signals from our brains to provide neurological data. EEG measures brain activity through electric currents. Electrodes on the scalp detect brain neural activity to measure and record brain activity.

EEG signals are often used to identify emotions. Over the past decade, several studies have examined the relationship between EEG waves and emotions. EEG measures brain responses to emotional stimuli non-invasively, quickly, and affordably [1]. As technology becomes part of our daily lives, EEG data for emotion classification can help create personalized apps that dynamically adapt to users' changing emotions and provide a more personalized and responsive user experience. With freely available EEG datasets like EEG Brain Wave Data Set: Feeling Emotions [2], [3], [4] EEG-based emotion analysis research is advancing.

As the scientific community seeks to create major emotional exchanges between humans and machines, reliable and implementable methods for recognizing human emotions are needed. [5]. Many studies have utilized machine learning and deep learning to classify EEG signals because they can recognize and learn from complex patterns. Classification techniques anticipate data classes, including assigning emotions to EEG signals. Our investigations are based on several machine learning and deep learning research on EEG-based emotion classification. Multiple machine learning-based techniques, such as Multi-Layer Perceptron (MLP) [1], [2], [6], K-Nearest Neighbor (K-NN) [6], [7], [8], [9] and Support Vector Machine (SVM) [2], [6], [8], [9] For EEG classification. In the research performed by [6] MLP has the best accuracy of 96.53%, KNN at 94.70%, and SVM at 96.90%. Various deep learning techniques, such as Convolutional Neural Networks[1], [9], [10], [11] (CNN), Long Short-Term Memory (LSTM) [12], [13], and Gated Recurrent Unit (GRU) [1], [12] are also used in EEG classification and provide excellent results. In the research performed by [1] CNN had the greatest average accuracy of 98.13%, GRU at 97.19%, and LSTM with 97.42 accuracy.

Our research utilized machine learning methods such as MLP, KNN, and SVM, as well as deep learning techniques including CNN, LSTM, and GRU. The differences from earlier research are seen in our different network architectures, parameter configurations, selection of activation functions, and utilization of cross-validation techniques. In addition, we employed grid search to choose the most suitable model by evaluating different combinations of parameters.

2. METHODS

An overview of the research methodology used in this research is illustrated in Figure 1. This research consists of five stages, from dataset preprocessing to performance evaluation, which will be explained in the sub-chapters. We can investigate a person's emotional reactions to specific surroundings by directly accessing their brainwave patterns [5]. Since emotions are encoded within chemical compositions that directly influence electrical brain activity, they may be classified using statistical aspects of the resulting brainwaves [4]. An electroencephalogram (EEG) is a test that measures and records electrical activity generated by the brain. EEG detects voltage variations that are caused by ionic current flows within brain neurons. In other words, it reads scalp electrical activity created by brain structures [14]. This research used a publicly accessible EEG dataset used for the emotion classification tasks. This dataset can be accessed online on the Kaggle website with the name EEG Brain Wave Data Set: Feeling Emotions [2], [3], [4], [13]. This dataset was collected using the Muse Headband sensors that have four electrode channels (AF7, AF8, TP9, and TP10). To minimize noise and preserve brainwave data, the Muse EEG headband uses a variety of artifact separation techniques. The dataset is taken from one male and one female participant for six minutes per emotion. The emotions of participants were evoked using six movie clip stimuli that represent positive and negative emotions. The dataset consists of 2132 samples 2548 input columns and 1 label column that indicates the emotion. There are 716 samples from the dataset labeled 'NEUTRAL', 708 are labeled 'NEGATIVE', and 708 are 'POSITIVE'.

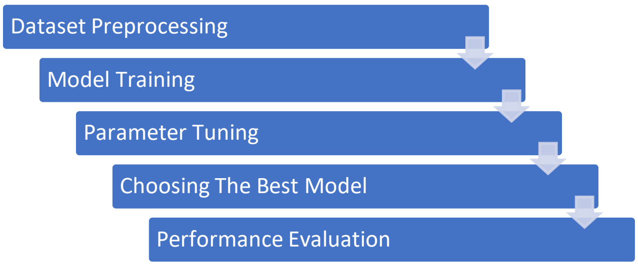


Figure 1. Research methodology overview.

2.1 Dataset Preprocessing

Initially, the data pre-processing step split the dataset into the input data (2548 × 2132 samples) and label data (1 × 2132 samples). The input data is then scaled using a standard scaler so that all the inputs are in the same data range. The standard scaler is a method that normalizes features by removing the mean and scaling them to have a variance of one. This implies that it alters the data to ensure the final distribution of each characteristic will have an average of 0 and a standard deviation of 1. The scaled input data was then divided into train and test sets at a ratio of 80:20. The last process in preprocessing is reshaping the data so that the data can be as a compatible input for CNN.

2.2 Model Training

To assess machine learning algorithms in robust situations, 5-fold cross-validation was applied. The training phase used 80% of the data, while the remaining 20% was used for testing.

2.2.1 Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP) is a sort of supervised artificial neural network that is made up of an input layer, at least one hidden layer, and an output layer. Backpropagation is used by MLP during the training phase. The MLP architecture used in this research is depicted in Figure 2. In this research, several MLP parameter combinations were examined to find the best ones. One buried layer has 50 neurons, two have 50 and 25, and three have 100, 50, and 25. The activation functions were ReLU and Tanh, while the solvers were Adam and SGD. Learning rates range from 0.0001 to 0.01. The learning rate might be constant, in scaling, or adjustable. Iterations are limited to 300.

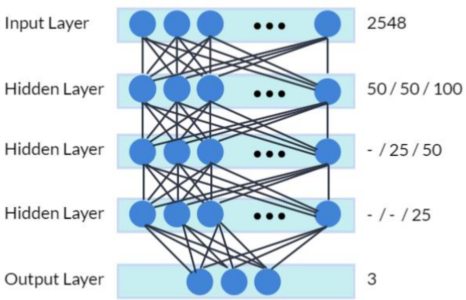


Figure 2. MLP architecture visualization showing the input, hidden, and output layers for EEG-based emotion classification

2.2.2 K-Nearest Neighbor (K-NN)

K-NN assumes that similar items have small distances or are close to each other. The value of K represents the number of nearest neighbors to be considered in predicting a class of new data. This research tested several K-NN parameter combinations to determine the best. Odd neighbor values from 3 to 30 were utilized; Manhattan and Euclidean distance metrics were utilized in this experiment, along with uniform and distance weights.

2.2.3 Support Vector Machine (SVM)

SVM operates by finding the optimal hyperplanes that separate data into different classes. The kernel function and penalty coefficient parameters have an impact on SVM performance; for this reason, improving the parameters added to SVM classifiers is crucial [15]. This research investigated several combinations of SVM parameters in an attempt to find the best combinations. C values in SVM varied between 0.1, 1, 10, and 100; Four types of kernels used were linear, RBF, poly, and sigmoid. The gamma values used were scale, auto, 0.1, 1, and 10.

2.2.1 Convolutional Neural Network (CNN)

CNN has convolutional layers that apply convolution operations to input data using filters to extract feature maps. Hyperparameters like the number of convolutional layers, size, and number of kernels, as well as pooling windows, have a significant impact on CNN performance [15]. Figure 3 shows the architecture of CNN that was used in this research. Several CNN parameter combinations were tested in this research to determine the optimal parameter combinations.

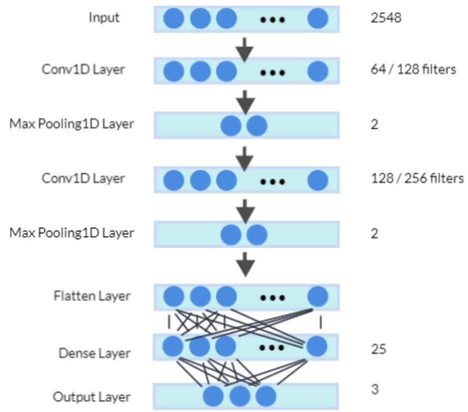


Figure 3. CNN architecture visualization for EEG-based emotion classification
Showing the input, convolutional layers, pooling layers, and fully connected layers.

First and second Conv1D layers employed filters between 64 and 128 and 128 and 256, respectively. Next to Conv1D is a max pooling layer with pool size 2. One flat layer precedes the thick layer. The dense layer has 25 neurons. Relu activation was used in the Conv1D and dense layers. The overall epoch was 15–25. The output (dense) layer used softmax activation. The Adam optimizer was used with learning rates of 0.001 and 0.01 for each combination.

2.2.2 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of ANN that consists of several Recurrent Neural Networks (RNN) that will make predictions based on previous states. The network's long-term dependencies are maintained by its gating mechanisms. Based on the gating mechanism, memory in the network can be released or stored on demand. The basic three components of an LSTM cell are gates. The first component is the forget gate, the second is the input gate, and the third is the output gate [12]. Figure 4 shows the LSTM architecture used in this research. In this research, several combinations of LSTM parameters were used, so that we can find the best parameter combinations. The variations of 64 and 128 for total units were used in the first LSTM layer, and 16 total units were used in the second LSTM layer; variations of 0.001 and 0.01 were used for the learning rate, and 15 and 25 were used for the total epoch. The activation function sigmoid was used in the LSTM layers. The first dense layer used the ReLU activation function with a total of neurons of 25, and in the output layer, the softmax activation function was used with a total of 3 neurons. Adam optimizer was used for all the combinations.

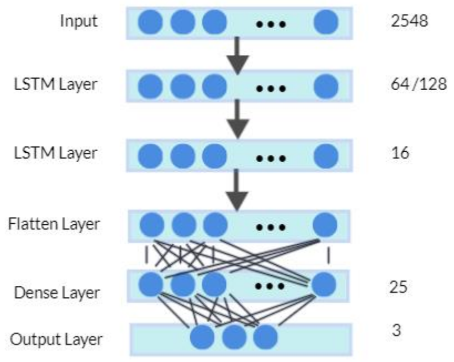


Figure 4. LSTM architecture visualization for EEG-based emotion classification Showing the input, LSTM layers, and fully connected layers.

2.2.3 Gated Recurrent Unit (GRU)

GRU is a newer form of RNN that is used in similar tasks as LSTM. GRU solves the classic RNN problem by combining two gates: the update gate and the reset gate. The gating mechanism in GRU allows it to update and reset its hidden state selectively. Portrayed in Figure 5 is the GRU architecture used in this research. This study examined many GRU parameter combinations to find the best. In the first GRU layer, the learning rate was 0.001 and 0.01; the total epoch was 15 and 25; the total units were 64 and 128, and the total units were 16 in the second. The GRU layer used sigmoid activation. The first dense layer has 25 neurons and ReLu activation. Softmax activation was used in the 3-neuron output layer. Adam optimizer ran all combinations.

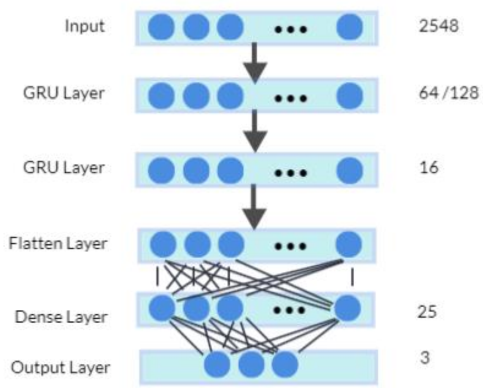


Figure 5. GRU architecture visualization for EEG-based emotion classification Showing the input, GRU layers, and fully connected layers.

2.3 Parameter Tuning

The grid search method was used to assess the performance of every potential combination. This was done using 5-fold cross-validation in machine learning, enabling us to determine the set of parameters that yielded the highest performance. Through careful and precise adjustments of these parameters, our objective was to improve the precision and resilience of our models in categorizing EEG data according to emotional states.

2.4 Choosing the Best Model

The model with the best parameter combination and accuracy value will then be selected to classify the test sets. A model that shows high accuracy and durability throughout the learning phase, as demonstrated by approaches like cross-validation, is more likely to effectively generalize to new and unexplored data. Consequently, the acquired patterns and correlations are transferrable to additional datasets beyond the one used for training.

2.5 Performance Evaluation

Testing results will be presented with a confusion matrix, accuracy, precision, recall, and F1-Score values. A comparison of four performance evaluation metrics values was conducted to assess the classification performance of the six classifiers. Calculations were made using the confusion matrix to get the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. Out of the four values indicated above, four metric values were finally obtained: accuracy, precision, recall, and f1-score. Equations (1), (2), (3), and (4) were used to calculate accuracy, precision, recall, and f1-score respectively[16].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{4}$$

To measure the classifier's performance, the classification scores were averaged over iterations. Accuracy and loss data during the training phase are monitored to track how well the deep learning classifier is learning from the training data and adapting to the underlying patterns. While loss measures the variation between expected and actual values, accuracy offers insights into the overall correctness of classification results.

3. RESULTS AND DISCUSSION

3.1 Dataset Preprocessing

The initial step in preprocessing the EEG data is to separate the labels from the data. Data and its label (positive, neutral, and negative) are shown in Figure 6. This separation guarantees that the labels (emotional states) and the features (EEG readings) can be processed independently. Next, the StandardScaler is employed to scale the EEG data. This scaler standardizes the features by removing the mean and scaling to unit variance. In essence, this phase is essential for guaranteeing that each feature contributes equally to the model's learning process by adjusting the data to have a mean of 0 and a standard deviation of 1.

	# mean_0_a	mean_1_a	mean_2_a	mean_3_a	mean_4_a	mean_d_0_a	mean_d_1_a	...	fft_742_b	fft_743_b	fft_744_b	fft_745_b	fft_746_b	fft_747_b	fft_748_b	fft_749_b	label
0	4.620	30.3	-356.0	15.60	26.3	1.070	0.411	...	20.300	20.300	23.50	-215.0	280.00	-162.00	-162.00	280.00	NEGATIVE
1	28.800	33.1	32.0	25.80	22.8	6.550	1.680	...	-21.800	-21.800	-23.30	182.0	2.57	-31.60	-31.60	2.57	NEUTRAL
2	8.900	29.4	-416.0	16.70	23.7	79.900	3.360	...	-233.000	-233.000	462.00	-267.0	281.00	-148.00	-148.00	281.00	POSITIVE
3	14.900	31.6	-143.0	19.80	24.3	-0.584	-0.284	...	-243.000	-243.000	299.00	132.0	-12.40	9.53	9.53	-12.40	POSITIVE
4	28.300	31.3	45.2	27.30	24.5	34.800	-5.790	...	38.100	38.100	12.00	119.0	-17.60	23.90	23.90	-17.60	NEUTRAL
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2127	32.400	32.2	32.2	30.80	23.4	1.640	-2.030	...	0.218	0.218	-21.70	95.2	-19.90	47.20	47.20	-19.90	NEUTRAL
2128	16.300	31.3	-284.0	14.30	23.9	4.200	1.090	...	-324.000	-324.000	594.00	-35.5	142.00	-59.80	-59.80	142.00	POSITIVE
2129	-0.547	28.3	-259.0	15.80	26.7	9.080	6.900	...	-160.000	-160.000	370.00	408.0	-169.00	-10.50	-10.50	-169.00	NEGATIVE
2130	16.800	19.9	-288.0	8.34	26.0	2.460	1.580	...	-27.600	-27.600	124.00	-656.0	552.00	-271.00	-271.00	552.00	NEGATIVE
2131	27.000	32.0	31.8	25.00	28.9	4.990	1.950	...	1.810	1.810	1.95	110.0	-6.71	22.80	22.80	-6.71	NEUTRAL

Figure 6. Initial data and its label; positive, neutral, and negative.

Subsequently, a LabelEncoder is employed to convert the labels into numerical values. The original categorical classifications (negative, neutral, and positive) are converted into numerical values by this encoding: 0 for negative, 1 for neutral, and 2 for positive values. This stage is crucial because numerical input is typically required to process machine learning algorithms.

The dataset is divided into training and testing sets using an 80-20 division following encoding. This implies that the model is trained using 80% of the data, while the remaining 20% is used to evaluate its performance. The last process in preprocessing includes the reshaping of the data to be compatible with a Convolutional Neural Network (CNN). CNNs typically necessitate input data that is of a particular shape, which includes a channel dimension. As a consequence, the training and testing data are expanded to include an additional dimension, resulting in shapes of (1705, 2548, 1) for the training data and (427, 2548, 1) for the testing data. This new shape is essential for CNN input, as it introduces a single channel dimension.

3.2 Model Training

In the current research, different learning algorithms were applied to emotion detection to improve classification accuracy. As was covered in the previous part, the EEG-based emotion-evoked signals are classified into positive, negative, and neutral emotions using six distinct models: machine learning-based models (MLP, K-NN, and SVM) which are implemented with scikit-learn, as well as deep learning-based models (CNN, LSTM, and GRU) which are implemented with TensorFlow.

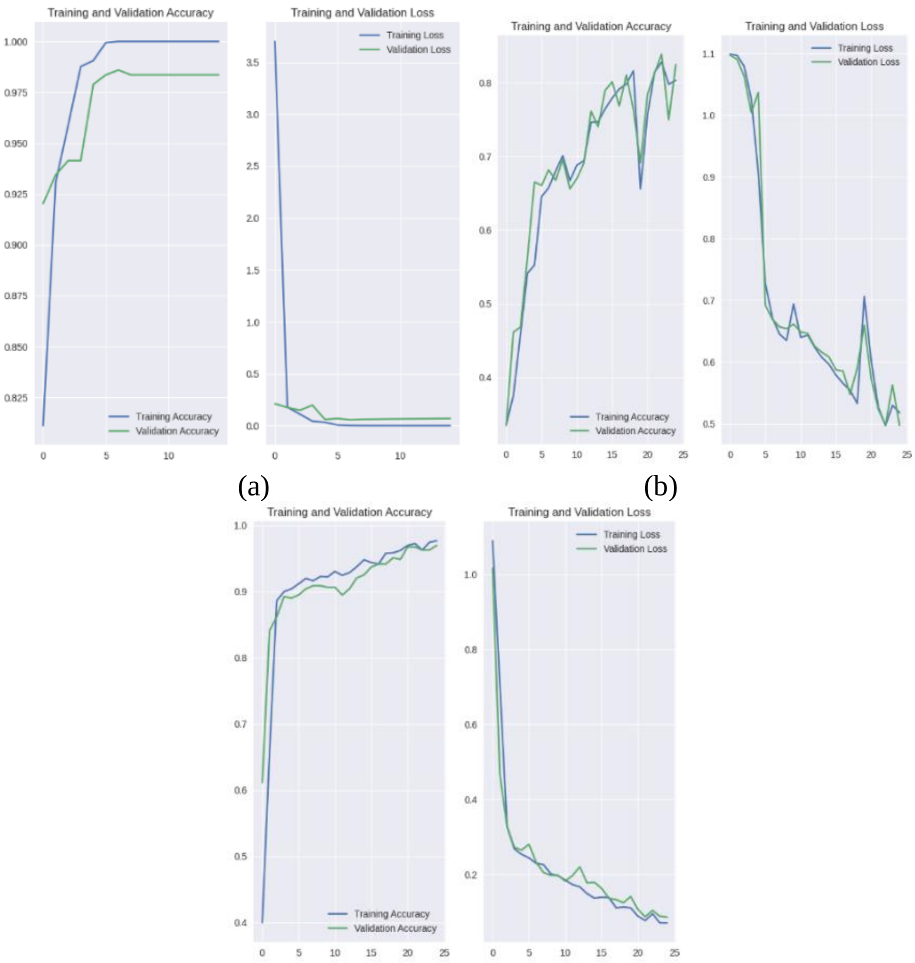


Figure 7. Training and validation metrics, showing training and validation accuracy and loss for CNN (a), LSTM (b), and GRU (c)

During CNN, LSTM, and GRU training and validation, we closely monitored the accuracy and loss metrics of our emotion categorization models. These measurements' graphical displays shown in Figure 7 highlight different patterns among the architectures. During training and validation, the CNN model demonstrated exceptional performance, attaining maximum accuracy and minimum loss. The GRU model produced excellent outcomes with competitive accuracy and loss metrics. On the other hand, compared to CNN and GRU, the LSTM model showed a little poorer accuracy and a higher loss despite its effectiveness. These visual aids highlight CNN's ability to identify complex patterns in EEG data, and GRU's strong performance confirms CNN's leadership in our research.

3.3 Parameter Tuning

A GridSearchCV-based MLP, KNN, and SVM classifiers were implemented in this stage. A parameter grid is established to investigate a variety of configurations as stated in subsection 2.2, such as the size of the hidden layer, the activation function, the solver, the regularisation strength, the learning rate, and the maximum iteration limit. GridSearchCV is implemented to conduct an exhaustive search of the specified parameter grid using cross-validation, to optimize accuracy. The model is subsequently adapted to the training data, and the most accurate parameters and their corresponding accuracy are printed.

Deep learning models, CNN, LSTM, and GRU, are implemented with different configurations as stated in subsection 2.2 by iterating over combinations of units, learning rates, and epochs. For each combination, a unique deep-learning model is created and trained. During training, accuracy and loss metrics for both training and validation data are tracked and stored. The results for each model configuration, including the unit count, learning rate, and epoch number, are printed for easy comparison.

3.4 Choosing The Best Model

Table 1 shows the optimal MLP, K-NN, and SVM parameter combinations from training and parameter selection. Next, testing data was classified using the best model and parameter combinations. We found that SVM is better than other algorithms in classifying positive, negative, and neutral emotions with a 0.98 accuracy rate. MLP and K-NN have 0.97 and 0.95 accuracy.

Table 1. Best parameter combinations for MLP, K-NN, and SVM

MLP		**K-NN**		**SVM**	
Parameters	Best Value	Parameters	Best Value	Parameters	Best Value
Activation Function	Tanh	Neighbors	3	C	100
Learning Rate	0.0001, constant	Distance	Manhattan	Kernel	RBF
Hidden Layer	(50, 25)	Weights	Uniform	Gamma	Scale

The optimal parameter combinations for CNN, LSTM, and GRU are produced by the training and parameter selection procedure and are displayed in Table 2. Testing data was then classified using the model with the best parameter combinations. According to our research, CNN fared better than other deep learning algorithms in terms of classification accuracy, achieving a 0.98 accuracy rate in identifying between neutral, positive, and negative feelings. While GRU and LSTM had accuracy rates of 0.97 and 0.82, respectively.

Table 2. Best parameter combinations for CNN, LSTM, and GRU

CNN		**LSTM**		**GRU**	
Parameters	Best Value	Parameters	Best Value	Parameters	Best Value

Epoch	15	Epoch	25	Epoch	25
Learning Rate	0.01	Learning Rate	0.01	Learning Rate	0.01
Filters in the 1st Conv1D Layer	128	Units in the 1st LSTM Layer	128	Units in the 1st GRU Layer	128
Filters in the 2nd Conv1D Layer	256	Units in the 2nd LSTM Layer	16	Units in the 2nd GRU Layer	16

3.5 Performance Evaluation

Table 3 shows macro averages for precision, recall, F1-Score, and MLP, K-NN, and SVM accuracy. The accuracy of our SVM model was 0.98, indicating its ability to recognize good cases. The SVM model's recall, or sensitivity, was 0.98, indicating its ability to detect true positive occurrences. The F1 score showed the model's balanced performance as 0.98 from the harmonic mean of accuracy and recall. SVM's overall prediction accuracy was 0.98, proving its ability to categorize instances in all classes.

Table 3. Performance metrics overview showing precision, recall, F1-Score, and accuracy values produced by the best model of MLP, K-NN, and SVM.

	Precision	**Recall**	**F1-Score**	**Accuracy**
MLP	0.97	0.97	0.97	0.97
K-NN	0.95	0.95	0.95	0.95
SVM	**0.98**	**0.98**	**0.98**	**0.98**

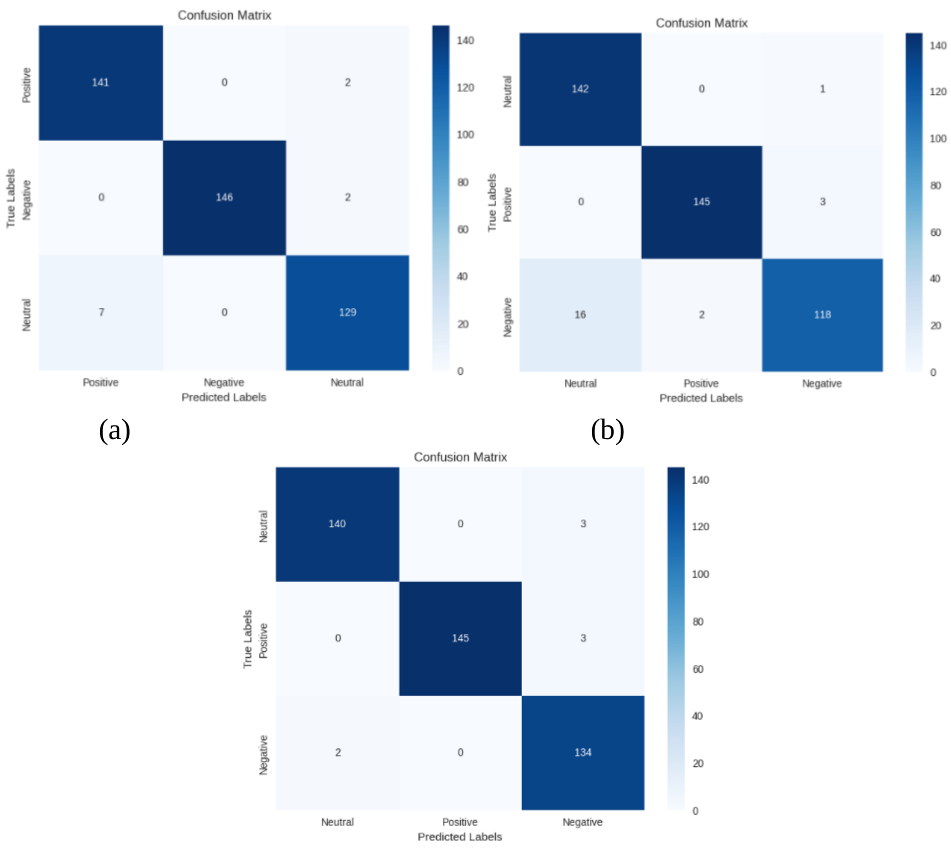


Figure 8. Confusion matrix visualization for MLP (a), K-NN (b), and SVM (c) showing classification results for every class.

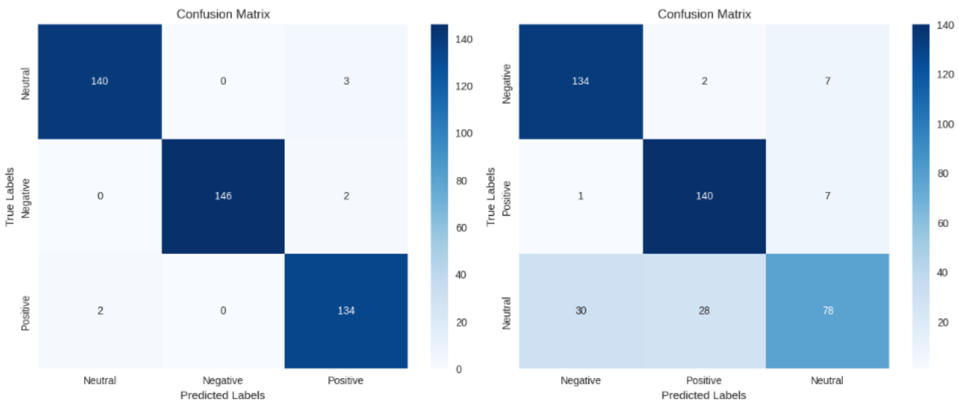
MLP, K-NN, and SVM confusion matrix in Figure 8 showed that our MLP model was good at classifying positive and negative emotions. The results of classifying neutral emotions by MLP are not as good as classifying negative and positive emotions; 7 neutral data were misclassified as positive data. Comparably, the K-NN confusion matrix revealed that this technique was not good enough in classifying negative emotions, compared with other emotions. Sixteen negative emotions were misclassified into neutral and 2 were misclassified into positive. This clarifies the possible drawbacks of the algorithm for classifying negative emotions. Furthermore, the SVM confusion matrix offered information about good classification results in every emotion, revealing the model's ability to identify patterns in the emotional data.

Table 4 displays the accuracy numbers for CNN, LSTM, and GRU and the macro averages for precision, recall, and F1-Score value. The precision of our CNN model, which tells us how well positive predictions performed, came in at 0.98, indicating that it can accurately identify positive examples. Moreover, the recall, or sensitivity, of the CNN model also demonstrated a 0.98 value, demonstrating its effectiveness in identifying real positive occurrences. As indicated by the F1 score, the model's balanced performance was demonstrated by the value of 0.98 derived from the harmonic mean of precision and recall. Furthermore, we were able to predict with an overall CNN accuracy of 0.98, indicating that the model is capable of reliably classifying cases across all classes.

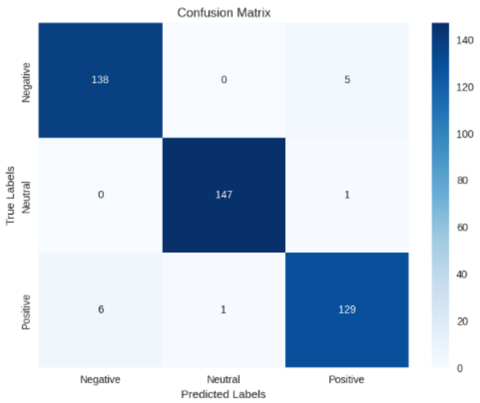
Table 2. Performance metrics overview showing precision, recall, F1-Score, and accuracy values produced by the best model of CNN, LSTM, and GRU.

	Precision	**Recall**	**F1-Score**	**Accuracy**
CNN	**0.98**	**0.98**	**0.98**	**0.98**
LSTM	0.83	0.82	0.81	0.82
GRU	0.97	0.97	0.97	0.97

The accuracy and precision of our emotion categorization models are greatly impacted by the machine learning and deep learning approaches that we choose. The strong machine-learning performance of the Support Vector Machine (SVM) confirms its effectiveness in managing the complex patterns seen in EEG data. Convolutional Neural Networks (CNNs) are a prominent performer in deep learning, demonstrating the capacity to extract intricate spatial information from neural input and deliver higher levels of accuracy. On the one hand, K-Nearest Neighbors (K-NN) struggle to classify negative emotions, whereas Long Short-Term Memory (LSTM) struggles to identify neutral emotional states.



(a) (b)



(c)

Figure 9. Confusion matrix visualization for CNN (a), LSTM (b), and GRU (c) shows classification results for every class.

Across all emotion classes, CNN showed the best accuracy and precision, making it the most reliable classifier. Although GRU performed admirably, it showed a notable misclassification of 6 positive instances as negative and 1 neutral instance. Conversely, although LSTM was able to accurately identify 78 neutral cases, it showed a significant difficulty in differentiating between neutral, negative, and positive emotions. It notably misclassified 30 neutral instances as negative and 20 as positive, pointing to the model's limitations. These insights from the confusion matrices shown in Figure 9 show that CNN is the most adept at identifying the complex patterns linked to various emotional states.

The research illustrates that the implementation of machine learning and deep learning methods will effectively categorize EEG signals to identify emotional states. The high precision scores, especially those achieved by the SVM and CNN models, demonstrate the promise of these models for practical use in emotion-aware human-computer interaction systems. This study adds to the expanding research on EEG-based emotion classification, providing valuable information on the selection of models and fine-tuning of parameters to enhance the accuracy of classification.

4. CONCLUSIONS

We investigated machine learning and deep learning techniques for categorizing emotions based on EEG data. The Support Vector Machine (SVM), K-Nearest Neighbours (K-NN), and Multilayer Perceptron (MLP) achieved accuracy ratings of 0.98, 0.95, and 0.97, respectively, in detecting emotional states. The deep learning techniques Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) all achieved high accuracy in identifying emotional states, with scores of 0.98, 0.82, and 0.97 respectively.

Overall, the SVM and CNN models classified EEG-based emotional states well, both with 0.98 accuracy. Deep learning models performed similarly to typical machine learning models in classifying emotional states from EEG data, according to the research. This research shows that advanced deep learning techniques can be used with classic machine learning approaches to classify emotions from EEG data. These advances in emotion identification systems may benefit neurofeedback, mental health monitoring, and human-computer interface systems.

REFERENCES

[1] N. Kumari, S. Anwar, and V. Bhattacharjee, “A Comparative Analysis of Machine and Deep Learning Techniques for EEG Evoked Emotion Classification,” Wireless Personal Communications. Accessed: Dec. 13, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s11277-022-10076-7

[2] J. J. Bird, L. J. Manso, E. P. Ribeiro, A. Ekart, and D. R. Faria, “A Study on Mental State Classification using EEG-based Brain-Machine Interface,” 9th International Conference on Intelligent Systems 2018: Theory, Research and Innovation in Applications, IS 2018 - Proceedings. Accessed: Dec. 13, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/8710576/

[3] Jordan J. Bird, “EEG Brainwave Dataset: Feeling Emotions.” Accessed: Dec. 13, 2023. [Online]. Available: https://www.kaggle.com/datasets/birdy654/eeg-brainwave-dataset-feeling-emotions

[4] J. J. Bird, A. Ekart, and D. R. Faria, “Mental Emotional Sentiment Classification with an EEG-based Brain-machine Interface HANDLE Project (EU FP7) View project EMG-controlled 3D Printed Prosthetic Hand for Academia View project,” Proceedings of theInternational Conference on Digital Image and Signal Processing (DISP’19). Accessed: Dec. 13, 2023. [Online]. Available: http://jordanjamesbird.com/publications/Mental-Emotional-Sentiment-Classification-with-an-EEG-based-Brain-machine-Interface.pdf

[5] N. S. Suhaimi, J. Mountstephens, and J. Teo, “EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities,” Computational Intelligence and Neuroscience. Accessed: Dec. 12, 2023. [Online]. Available: https://www.hindawi.com/journals/cin/2020/8875426/

[6] F. K. Bardak, M. N. Seyman, and F. Temurtaş, “EEG Based Emotion Prediction with Neural Network Models,” Tehnicki Glasnik. Accessed: Dec. 10, 2023. [Online]. Available: https://hrcak.srce.hr/283785

[7] Shashank Joshi and Falak Joshi, “Human Emotion Classification based on EEG Signals Using Recurrent Neural Network And KNN,” International Journal of Next-Generation Computing. Accessed: Dec. 15, 2023. [Online]. Available: https://ijngc.perpetualinnovation.net/index.php/ijngc/article/view/691

[8] A. S. M. Miah, J. Shin, M. M. Islam, Abdullah, and M. K. I. Molla, “Natural Human Emotion Recognition Based on Various Mixed Reality(MR) Games and Electroencephalography (EEG) Signals,” 5th IEEE Eurasian Conference on Educational Innovation 2022, ECEI 2022. Accessed: Dec. 13, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9829482

[9] A. A. Rahman *et al.*, “Detection of Mental State from EEG Signal Data: An Investigation with Machine Learning Classifiers,” KST 2022 - 2022 14th International Conference on Knowledge and Smart Technology. Accessed: Dec. 15, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9729084

[10] R. Qiao, C. Qing, T. Zhang, X. Xing, and X. Xu, “A novel deep-learning based framework for multi-subject emotion recognition,” in *ICCSS 2017 - 2017 International Conference on Information, Cybernetics, and Computational Social Systems*, 2017. doi: 10.1109/ICCSS.2017.8091408.

[11] Z. Jiao, X. Gao, Y. Wang, J. Li, and H. Xu, “Deep Convolutional Neural Networks for mental load classification based on EEG data,” Pattern Recognition. Accessed: Dec. 12, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0031320317304879

[12] M. K. Chowdary, J. Anitha, and D. J. Hemanth, “Emotion Recognition from EEG Signals Using Recurrent Neural Networks,” Electronics (Switzerland). Accessed: Dec. 14, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/11/15/2387

[13] J. J. Bird, D. R. Faria, L. J. Manso, A. Ekárt, and C. D. Buckingham, “A deep evolutionary approach to bioinspired classifier optimization for brain-machine interaction,” Complexity. Accessed: Dec. 13, 2023. [Online]. Available: https://dx.doi.org/10.1155/2019/4316548

[14] S. M. Alarcão and M. J. Fonseca, “Emotions recognition using EEG signals: A survey,” IEEE Transactions on Affective Computing. Accessed: Dec. 12, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/7946165

[15] M. Saeidi *et al.*, “Neural decoding of EEG signals with machine learning: A systematic review,” Brain Sciences. Accessed: Dec. 14, 2023. [Online]. Available: https://www.mdpi.com/2076-3425/11/11/1525

[16] D. Jung, J. Choi, J. Kim, S. Cho, and S. Han, “EEG-Based Identification of Emotional Neural State Evoked by Virtual Environment Interaction,” International Journal of Environmental Research and Public Health. Accessed: Dec. 18, 2023. [Online]. Available: https://www.mdpi.com/1660-4601/19/4/2158